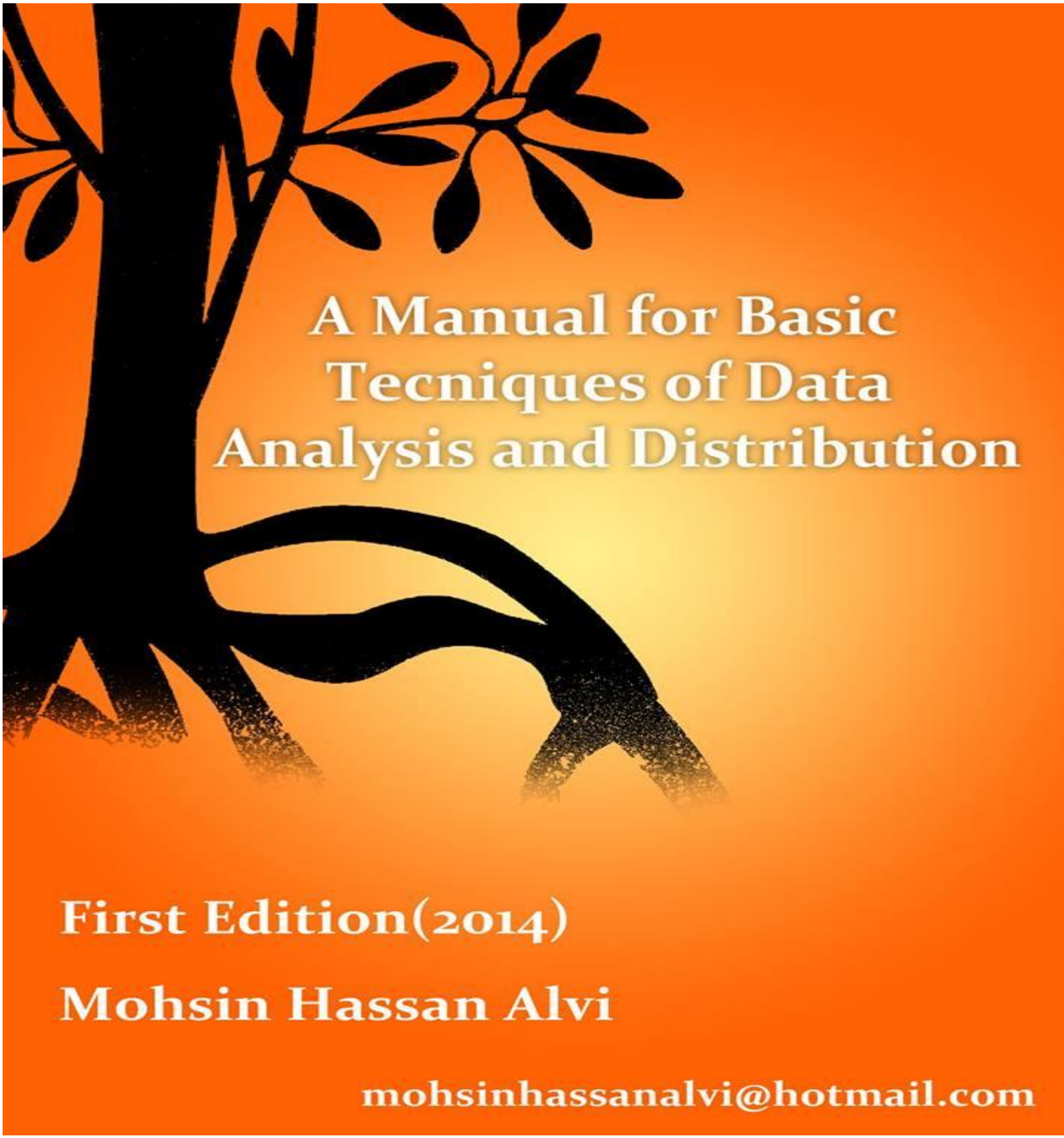
A Manual for Basic Techniques of Data Analysis and Distribution

Alvi, Mohsin

Iqra University, Pakistan Institute of Learning and Living

23 November 2014


A Manual for Basic
Tecniques of Data
Analysis and Distribution
First Edition(2014)
Mohsin Hassan Alvi
mohsinhassanalvi@hotmail.com

flyleaf - 1

This page intentionally left blank

A Manual for Basic Techniques of Data Analysis and Distribution (2014)

Written By

Mohsin Hassan Alvi
Karachi, Sindh, Pakistan
Pakistan Institute of Learning and Living
+92-334-3833358, +92-345-2295567
mohsinhassanalvi@hotmail.com

Acknowledgement

Many Thanks to Almighty God, who is Merciful and Beneficent

I would like to express my gratitude to the many people who saw me through this manual book; to all those who provided support, talked things over, read, wrote, offered comments, allowed me to quote their remarks and assisted in the editing, proofreading and design.

I would like to thank Dr. M. I. Subhani for helping me to write this manual. I also would like to thank my Parents, Teachers, Friends and the Rest of my family, who supported and encouraged me in spite of all the time it took me away from them. It was a long and difficult journey for them.

Thanks to Publications E-Servers - without you, this manual book would never find its way to the Web and to so many people who cannot read the printed form of manual book.

Last and not least; I beg forgiveness of all those who have been with me throughout the preparation of manual preparation and whose names I fail to mention.

Abstract

A manual is designed to support and help the basic concepts of statistics and its implications in econometric, beside this, interpretation of further statistical techniques have been shown as well by illustrations and graphical methods. It is comprised on several instances of test, obtained from statistical software like SPSS, E-views, Stata and R-language with the understanding of their research models and essentials for the running the test. A basic of manual is included on two elements, first is Distributions of data and second is techniques involve in data analysis. The first part considered the thoughtful meaning of data nature in statistical aspects whereas the other important part explained tactics incorporate with different sets of data and also concept of assimilation of tests in a form of flow chart. The factors have been identified through tests are associations, causes, effects, impacts, relations and integrations that belong to time series sets of observations. On the other hand, a review has been performed of non-time series statistics by introducing techniques in the form of list of tests. The most of the theme of manual is designed in bullets form in order to provide better understanding to reader for the purpose of analyses. Beside these, supported studies of famous authors have been reviewed; hence, technical terminologies of synonyms have been demonstrated in glossary section.

CONTENTS

1. Econometrics..... 7
2. Distribution of Data..... 9
3. Transformation..... 11
4. Parametric & Non-parametric..... 11
5. Specification Error..... 12
6. Presence of Outlier..... 14
7. Sample Size..... 14
8. Co-linearity & Multi-co-linearity..... 15
9. Statement of Falsification..... 17
10. Techniques Involve in Data Analysis..... 18
11. Time Series Econometrics..... 18
12. ADF Unit Root Test..... 19
13. ARLD Model..... 24
14. AR & VAR..... 25
15. Unidirectional Causality..... 31
16. Bidirectional Causality..... 34
17. J-BOX Q-Statistics..... 35
18. Cross Correlation..... 41
19. Co-Integration..... 43
20. ECM..... 49
21. Endogenity..... 51
22. Exogenity..... 51
23. Sum up All Tests..... 53
24. Non-Time Series Econometrics..... 55
25. References..... 58
26 Synonyms Glossary..... 59

flyleaf - 2

This page intentionally left blank

ECONOMETRICS

- Econometrics is a branch of social science.
- Econometrics starts when statistics ends, and econometrics formed with the violation of statistics.
- Statistics does not tell exact interpretation of data where as econometrics tells the exact interpretation of data.
- Empirical testing is best in econometrics.

Software for data analysis:

SPSS

R-language

E-views

Stata

Types Axioms:

- Rationality
- Convexity
- Continuity
- Revealed preference
- Random or stochastic

Types of data:

Time series data: if data is explaining time by time period.

Panel data: data is defining with respect to period. Panel data itself a time series data.

Cross sectional data: data is defining in a one period of time.

Autocorrelation:

1) If we want to find autocorrelation in time series data then we use Durbin Watson test.
2) If we want to find autocorrelation in panel data then we use H-test.

   In panel data we do not take out autocorrelation of panel data in statistics but in econometrics we use H-test.

Distribution:

Data

Parametric

Shapiro test (tells distribution)

(if test is non-sig then normal distribution)

It's ok. Normal distribution is necessary in parametric data

(if test is sig then not normal distribution)

Transformation will be applied

For time series data: 1st and 2nd difference

For non-time series data: Arithmetic/Geometric

Non-parametric

K-Smirnov test (tells distribution)

(if test is non-sig then normal distribution)

Parametric test will be applied

(if test is sig then not normal distribution)

Non-parametric test will be applied

First we will find that data is parametric or non-parametric.

Observation of parametric data = 5kg, 25km/hrs, $60m^3$ etc (vector or unit is defined)

Observation of non-parametric ordinal data: 5,12,18,16,21 etc (vector is not defined)

For Non-parametric Data:

If data is non-parametric then we will apply k-Smirnov test. It will tell that data is normal distributed or not for non-parametric data. If value of test comes significant then data is not normal distributed and we can apply non-parametric test. And if value of test comes non-significant then data is normal distributed and we can apply parametric test.

For Parametric Data:

If data is non-parametric then we will apply Shapiro test. It will tell that data is normal distributed or not for parametric data. If value of test comes non-significant then data is normal distributed. If data is normal distributed then it's ok and we will do nothing because normal distributed is necessary in parametric data. And if value of test comes significant then data is not normal distributed, so we will make data normal distributed by transformation method. If data is time series then we will apply arithmetic and geometric transformation for making data normal. If data is not time series then we will apply 1st difference and 2nd difference.

For distribution, first we will analyze data that is present in parametric form or non-parametric form. If data is parametric then normal distribution is necessary. If data is not normally distributed then we have to make data normal by using transformation. But if data is non-parametric then we will use K-Smirnov test. It will tell that data is normal or not but beside this it also tells that weather we will apply parametric test or non-parametric test.

If you are predicting something then your data will be normal.

If you are finding cause & affect then you data will be not normal.

Methodology will decide about our data that it is normal or not.

- For regression both variable are in normal distribution.
- For blues only dependent variable in normal distribution.
- For censored regression, dependent variable are non-parametric or semi parametric. We censored dependent variable so we censored its distribution.
- If we want to check shock/non-stationarity/oscillation/bombs in time series then no normal distribution data is required.
- If data found in normal distribution then we do spectral analysis/ de-trending by using trigonometric transformation in order to find shocks.

Transformation:

For making data in normal distribution we do transformation.

1) For making data normal in Non-time series data we use

   Arithmetic = square root / reciprocal

   Geometric = log

2) For making data normal in time series data we use

   1st difference and 2nd difference

   If any case 1st difference and 2nd difference is not able to investigate that means sure cannot be investigated.

3) For making data not normal in time series data we use

   Trigonometric transformation

   De-trending or Spectral analysis

4) For making data not normal in non-time series data we do not need to do anything because it would be already present in the form of not normal distribution.

Parametric & non-parametric:

- There is no need to check distribution of non-parametric data in statistics.
- We can check distribution of non-parametric data in econometrics.
- For parametric data we use Shapiro test.

  If its value comes significant then distribution is not normal.
- For non-parametric data we use keys 7 rock test.

  If its value comes significant then distribution is not normal.

- If answer comes in not normal distribution then we apply non-parametric test.
- If answer comes in normal distribution then we apply parametric test.

- If data is non-parametric nominal it means, it's a categorical data so we can't investigate its distribution.
- If data is non-parametric ordinary then we can investigate its distribution. We apply K-Smirnov test then we check that either we should apply parametric test or non-parametric test.

Specification Error:

Specification error can be identified by literature support not by econometrical test or econometrics.

Specification error is found in model, and sometimes in data as well.

1) If we add any irrelevant variable in model or we did not add any relevant variable in model then it become specification error.

   Eg: y = a + b1x1 + b2x2 + ET X3 is not included

   X2 is irrelevant or X3 is relevant

2) If single log is required and we have taken double or semi log, if double log is required and we have taken single or semi log, if semi log is required and we have taken single or double log.

   Eg: logy = a + b1logx1 + b2logx2 +ET

   Single log is required

3) If we formed only single equation but model required two equations or if we formed two equation and model required on single equation.

   Eg: y = a + bx

   Y = a + bx + ET

   Only one equation is required

4) If model should be OLS and you have made scaled OLS means you took wrong model.

   Eg: we have to introduce lag variable, mean we are finding out current GDP from previous GDP data and we have applied ADF model but ARLD model is required.

5) If the relationship is studied then model should not be deterministic.

   Eg: it should be stochastic because we are doing research in social science but not in pure science.

6) Model required variable with relative value while model has variable with absolute value.

   Eg: income in % = a +b (expenditure Rs25000)

   Expenditure should also in %

7) If co-linearity exists between both independent variables then we transform any one variable, not both. If we transform both variables then it becomes specification error.

8) Specification error includes taking variable in wrong form.

   Eg: house hold expense = a + b salary

   We should take house hold income instead of salary because it will define better relation.

9) Co-linearity is the specification error.

   Eg: both independent variables are correlated with each other.

10) If DV and IV are same then it is also specification error.

    Eg: expo is in front of civic center and civic center is in front of expo. So address can't be defined.

11) If you run wrong test then it also includes in specification error.

12) Presence of type 1 and type 2 errors in research is considered as specification error.

13) Data was required in normal distribution and you took not normal data or vice versa then it is specification error.

14) If model requires linear association and you have taken non-linear then its specification error.

Presence of Outlier:

- Outlier= observation of data which are not following a trend.
- Outliers disturb arithmetic mean.
- Z = more than 3.384 then observation is significant outlier.
- Grubbs test (which is in R language) helps to find out exact number of outlier in data by giving Z value of each observation.
- If we delete outliers there will be two issue:

1) Sample size would reduce.
2) Sample will become biased.

- There are two ways to get rid of outliers. One is to dissolve them or either eliminates them.
- For dissolving: We do log transformation for dissolving outliers. It will make the Z value of observation is 1. Something so due to this observation of outlier comes under normal observation.
- For eliminating: if outlier comes from human error so it should be eliminated.

Sample Size of Data:

- If data is non-parametric then population is infinite so sample size will be around 280 – 380.
- If data is parametric and not time series then sample size at least 30 observations.

Eg: 30 years, 30 countries, 30 months

- If data is parametric and time series then sample size should be 60 +
- Sample size and sample unit are different. (Sample size: 300 student, sample unit: 5 universities)

Co-linearity and Multi-co-linearity:

If relationship exist between two independent variable then co-linearity exist in each other.

Eg:

1) DV = a1 + b1 IV1 + b2 IV2 + ET if IV1 & IV2 have relation then co-linearity exist
2) DV = a1 + b1IV1 +b2IV2 + b3IV3 + ET if IV1, IV2 & IV3 have relation then multi-co-linearity exist

Due to co-linearity independent variables can't explain dependent variable completely so its an issue.

Co-linearity increases ET so:

Cause: presence of co-linearity

Effect: increase in ET

So if we have idea about effect means ET so we can conclude cause means co-linearity.

Indication of co-linearity:

Gardge magnitude: increase, (husky), around 1000

Gatrade magnitude: increase

Tolerance: decrease

Adjusted R square: non-significant, increase, more than 0.9

F value: non-significant, less than 3.84

VIF= more than 2

VIF is inversely proportional to tolerance.

Solution to fixed co-linearity:

Co-linearity in variables is the specification error.

There are 2 solutions to resolve co-linearity:

1) Transformation method:

- We do transformation to create dissimilarity, so we do transformation any one variable.
- If the issue of co-linearity is not resolve by transformation then we do double transformation.
- The original data and accurate information loss by transformation and research can be biased, so the best method is substitution.
- In substitution method, it removes co-linearity without transform the data.

2) Substitution method:

Y = a + b1x1 + b2x2 + ET

After run model, model will be:

Y = a + b1x1 + b2x2 ------ eq 1

If VIF value comes more than 2 then x1 and x2 are co-linear with each other.

Where, x1 = a + b3x2 -------eq 2

X2 = a + b4x1 ------- eq 3

Putting these two equations in eq 1 then add both, so we will:

Y = 0.5a5 + 0.5a7 + 0.5b6x2 + 0.5b8x1

So the value of b1 and b2 were same. Now both have change to b6 and b8. So co-linearity has resolved.

Statement of Falsification:

- A statement which denies, revamp or make it doubtful existing hypothesis, principle, theory and law is called statement of falsification.
- It also helps in extracting assumptions.
- If specification error present in model then hypothesis can be challenged through statement of falsification.
- If statement of falsification accepts then it cut it off the theory from root.

  If statement of falsification rejects then it revalidate of theory.

  Assumption can be formed in theory in order to get rid of from statement of falsification.
- How many assumptions are present in theory, it tells number of statement of falsification came against theory.

Reason of statement of falsification:

1) Specification error
2) Observation
3) Logical reasoning
4) Literature support
5) Violation of axiom
6) Presence of type errors

   Eg:

   Theory: Brand awareness explains intention to purchase.

   SOF: Brand equity does not explain intention to purchase.

Techniques Involve In Data Analysis:

1) If two trends are associated then we check cross correlation between them.
2) If two equations/expressions are associated then we check co-integration.
3) If two variables are associated then we check correlation.
4) If two observations are associated then we check auto correlation.
5) If two independent variables are associated then we check co-linearity.
6) If two dependent variables are associated then we check sure.
7) For transforming dependent variable we use blues.
8) For finding the non-correlated objects we use sure investigate.
9) Sure tells hidden relation.

Time Series Econometrics:

If data is time series then we apply time series econometrics.

Time series researches are found in large number as compare to non-time series researches.

Time series researches are mostly found in finance and economic and sometimes in marketing.

A thing should be appropriate in which you are using data.

Coefficients:

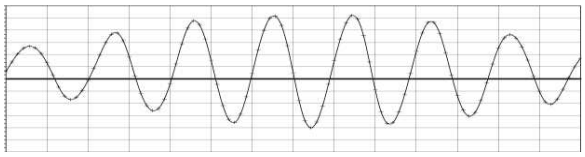
- The coefficient of regression is beta (β)
- The coefficient of correlation is the coefficient of correlation
- The coefficient of auto correlation is Rho
- The coefficient of auto regression is Rho
- The coefficient of co-integration is Log-Likelihood

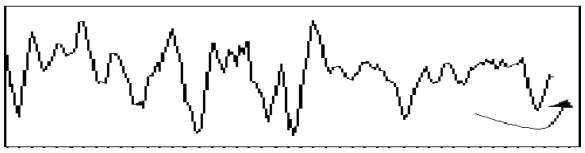
ADF-test/ADF unit root test:

ADF is the abbreviation of Augment Degree Fuller.

1) This test checks non-stationarity, shocks, bombs and oscillation in time series data.
2) Beside this it also tells that shocks are temporary or permanent.
3) It tells that is there any room to correct error term.
4) It tells that when we will apply ECM (Error Correction Model).

ADF test explains that is data in unit root or not means what is ADF t value and relation with Micron critical value. It tells comparison of ADF correspondence t value with MicKinnon critical value.


(Variation) having small wavelength


(Shocks) having large wavelength

Two questions:

1) Why shocks are present in data?

   Answer: Variation in data, means shocks in data come on the basis of time and place. Means observation changes time by time and place by place. If data is changing by time so it is time series data. So it is necessary in time series data that variation will come by time because time is changing automatically.

2) Why should we investigate them?

Answer: We have main concern about change of value neither absolute value that is why we analyze variation in data. Eg: We do not concern, what is share price of today but we have concern that what will be share price of tomorrow and the day after tomorrow.

In e-view software, Test critical is Micron critical value.

If ADF-test value is greater than Micron critical value then presence of non-stationarity.

If ADF-test value is less than Micron critical value then presence of stationarity.

ADF model is being made in OLS frame work.

Eg:	$\triangle Yt$ =	$aYt-1$ +	$x't$ +	$B1\triangle Yt-1$ +	ETt
Alpha value		-0.00963	1.562	0.38	
t-value		[-1.1337]	[1.4066]	[5.91420]	

Taken from Dr. Subhani's research: Information management and business review. Vol 2 no.6 pp.246-251, june 2011. (Monetary shocks or real shocks)

Notation:

a = root / absolute value of share

$\triangle$ = delta / change in share price

B = beta tells the change

X' = Exogenous Regressor = correction of error term and included in model

ET = error term

Yt – 1 = previous period observation

Findings:

1) ADF > MCV

   Presence of non-stationarity.

2) Co-efficient of augmented dickey fuller test is 1 or more

Shocks are permanent.

3) t-value of C is 1.5 or more (significant)

   Model has a room to correct error term and ECM may be applied to correct error term.

4) Confirmation test: prob or p-value is more than 0.05 (non-significant)

   Presence of non-stationarity.

Software information:

For current lag means zero order and "t"

Lag 1 = t-1 and Lag 2 = t-2

1) (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1. (run)
2) (Eliminate) year, date and month.
3) (Go to) Quick, series statistics, unit root test.
4) Insert column for consideration.
5) (Select) level, Schwarz info criterion, intercepts, maximum lag = 10.

Example#1:

Augmented Dickey-Fuller Unit Root Test on MARKET_1

Null Hypothesis: MARKET_1 has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.936347	0.7696
Test critical values:	1% level	-3.548208	
	5% level	-2.912631	
	10% level	-2.594027	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MARKET_1)
Method: Least Squares
Date: 11/20/11 Time: 17:18
Sample (adjusted): 3 60
Included observations: 58 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
MARKET_1(-1)	-0.004689	0.005008	-0.936347	0.3532
D(MARKET_1(-1))	0.337802	0.134788	2.506162	0.0152
C	123.2966	44.60893	2.763946	0.0078

R-squared	0.114775	Mean dependent var	132.8127
Adjusted R-squared	0.082585	S.D. dependent var	93.55527
S.E. of regression	89.60888	Akaike info criterion	11.87912
Sum squared resid	441636.4	Schwarz criterion	11.98570
Log likelihood	-341.4946	Hannan-Quinn criter.	11.92064
F-statistic	3.565562	Durbin-Watson stat	2.068954
Prob(F-statistic)	0.034992		

Findings:

1) ADF test value > MCV

   -0.93 > -2.59 , -2.91 , -3.54 at 1% , 5% , 10%

   Presence of non-stationarity

2) Coefficient of ADF = -0.004689 < 1

   Shocks are temporary

3) t-value of C = 2.763946 > 1.5 (significant)

   Model has a room to correct error term and ECM may be applied to correct error term.

4) Confirmation test: prob = 0.7696 (insignificant)

   Presence of non-stationarity

5) Model: $\triangle M1t = \alpha M1t\text{-}1 + \delta Xt + \triangle M1t\text{-}1 + ET$

-0.0046	123.2966	0.3378
[-0.936]	[2.763]	[2.506]

Example#2:

Augmented Dickey-Fuller Unit Root Test on D(TOTAL)

Null Hypothesis: D(TOTAL) has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic based on AIC, MAXLAG=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.145387	0.9662
Test critical values:	1% level	-3.565430	
	5% level	-2.919952	
	10% level	-2.597905	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TOTAL,2)
Method: Least Squares
Date: 11/20/11 Time: 18:26
Sample (adjusted): 10 60
Included observations: 51 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
D(TOTAL(-1))	0.007984	0.054916	0.145387	0.8851
D(TOTAL(-1),2)	-0.518730	0.152524	-3.400975	0.0015
D(TOTAL(-2),2)	-0.193619	0.166844	-1.160479	0.2524
D(TOTAL(-3),2)	0.076819	0.163786	0.469024	0.6415
D(TOTAL(-4),2)	-0.248211	0.162103	-1.531198	0.1332
D(TOTAL(-5),2)	-0.413428	0.173962	-2.376543	0.0221
D(TOTAL(-6),2)	-0.587868	0.182701	-3.217643	0.0025
D(TOTAL(-7),2)	-0.482658	0.180428	-2.675075	0.0106
C	853.3509	1764.574	0.483602	0.6312

R-squared	0.392705	Mean dependent var	400.5374
Adjusted R-squared	0.277030	S.D. dependent var	2092.560
S.E. of regression	1779.255	Akaike info criterion	17.96456
Sum squared resid	1.33E+08	Schwarz criterion	18.30547
Log likelihood	-449.0963	Hannan-Quinn criter.	18.09483
F-statistic	3.394891	Durbin-Watson stat	2.112628
Prob(F-statistic)	0.004301		

Findings:

1) ADF test value > MCV

   1.45 > -2.59 , -2.91 , -3.56 at 1% , 5% , 10%

   Presence of non-stationarity

2) Coefficient of ADF = 0.007984 < 1

   Shocks are temporary

3) t-value of C = 0.483602 < 1.5 (insignificant)

   Model does not has a room to correct error term.

4) Confirmation test: prob = 0.9662 (insignificant)

   Presence of non-stationarity

5) Model: $\triangle M1t = \alpha M1t\text{-}1 + \delta Xt + \triangle M1t\text{-}1 + ET$

-0.0046	123.2966	0.3378
[-0.936]	[2.763]	[2.506]

Auto Regressive Lag Distribution Model (ARLD Model):

In OLS or regression model, if lag variable of dependent variable is introduced then model become ARLD model. (OR) If lag variable is introduced in model and it regress dependent variable then model is ARLD.

Why we called its Auto Regressive Lag Distribution model?

Because lag variable is distributed in model and dependent variable is regress with its lag.

If dependent variable is regress with its lag then its mean that auto regression is present.

OLS model: $DV = \alpha + \beta IV + ET$

ARLD model: $DV = \alpha + \beta 1IV + \beta 2DVt\text{-}1 + ET$ therefore, DVt-1 is a lag variable

Current lag previous lag

Year	Xt	Xt-1
2011	10	20
2010	20	24
2009	24	23
2008	23	27
2007	27	29
2006	29	31
2005	31	30
2004	30	35

Xt-1 = lag variable and Xt+1 = log variable

Xt-1 = lag variable of previous year

Xt-2 = lag variable of previous to previous year

Xt-3 = lag variable of previous to previous to previous year

Vector Auto Regression (VAR) and Auto Regression (AR):

1) If Stationarity is present in one time series data then AR is applied.
2) If Stationarity is present in two or more time series data. In which first is dependent variable and second is independent variable then IV is predicting DV, so VAR is applied.

If non-stationarity is present in 2 or more non-time series data then we apply regression and OLS model.

ADF model must be applied before applying VAR & AR because ADF tells non-stationarity in time series data.

1) $DV = \alpha + \beta 1lag1DV + ET$ (ARLD for AR)
2) $DV = \alpha + \beta 1lag1DV + \beta 2lag2DV + \beta 3lag1IV + \beta 4lag2IV + ET$ (ARLD for VAR)

In both equation, model is ARDL and econometrics is VAR and AR

1) If DV is regress with its previous lag variable then ARLD model is for AR.
2) If DV is regress with its previous lag variable and as well as previous lag variable of other variable then ARLD model is for VAR.

1) For 1 time series = AR is applied
2) For 2 or more time series = VAR is applied

X, Y is absolute value

$\triangle X$, $\triangle Y$ is relative value

1) In AR we use absolute term

$X = \alpha + \beta 1lag1X + ET$

2) In VAR we use relatives & absolute term both

$X = \alpha + \beta 1lag1X + \beta 2lag2X + \beta 3lag1Y + \beta 4lag2Y + ET$

$\triangle X = \alpha + \beta 1lag1 \triangle X + \beta 2lag2 \triangle X + \beta 3lag1 \triangle Y + \beta 4lag2 \triangle Y + ET$

1) We apply AR in SPSS-13
2) We apply VAR in e-views

1) Example for AR:

In spss-13. (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1. (run)

Then analyze, time series, auto correlation

Select DV = market_1 & IV = market_2

Select Prais-winsten then OK

Auto regression = auto correlation

Durbin Watson = around 2 means no auto correlation

Durbin Watson = Less than 2 means positive auto correlation.

Durbin Watson = More than 2 means negative auto correlation.

Durbin Watson tells the sign of Rho. Weather it is positive or negative.

Rho tells the association between current lag & previous lag.

In regression β is coefficient and in auto regression Rho is coefficient.

Rho is a coefficient of lag1 DV.

Good model: Durbin Watson ~ Rho

ARLD is applied if auto correlation is present and Durbin Watson is significant.

If t-value comes 1.5 or more, then it's significant. That mean, its coefficient and its variable are also significant.

If Rho = zero then ARLD model can not apply. That mean, there is no coefficient of previous lag variable.

ARLD model can apply, if auto correlation is present and DW is significant.

In final Iteration DW must be less than 2.

Iteration 7

Autocorrelation Coefficient

Rho (AR1)	Std. Error
.944	.044

The Prais-Winsten estimation method is used.

Model Fit Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
.881	.776	.768	127.960	1.028

The Prais-Winsten estimation method is used.

ANOVA

	Sum of Squares	df	Mean Square
Regression	3234121	1	3234120.825
Residual	933308.7	57	16373.838

The Prais-Winsten estimation method is used.

Regression Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig
Subscribers for Market 2	.165	.012	.881	14.054	.000
(Constant)	2354.574	449.517		5.238	.000

The Prais-Winsten estimation method is used.

Findings:

1) Rho = 0.944. That means, if IV (increase in 1 unit) then DV (increase in0.944 units)
2) DW = 1.028. Less than 2. That mean, significant and there is positive autocorrelation b/w DV and its lag.
3) DW = 1.028. Less than 2. That mean, the sign of Rho is also positive. (i.e. Rho = +0.944)
4) Rho is not in ($\triangle$) delta form. So here it is showing absolute value, not relative value.
5) Model:

   DV = α + β1lag1DV + β2IV + ET

   M1 = α + β1lag1M1 + β2M2 + ET

Alpha (coefficient)	B=2354	Rho=0.944	B=0.165
t-value	(5.238)	(DW=1.028)	(14.054)

6) Intercept and market 2 are significant, having t-value significant. While previous lag of market 1 is insignificant.
7) Market 2 and intercept significantly predicting market 1.
8) Intercept is significant. That mean, there is some value of intercept. And if we make zero all other IV then there will be some value of DV.

2) Example for VAR:

VAR is present in absolute & relative term both.

Significant = predicting

[ ] = t-statistics value

( ) = standard error

Without bracket = coefficient

If t-value comes 1.5 or more, then it's significant. That mean, its coefficient and its variable are also significant.

If f-value come 3.84 or more, then it's significant. That mean, R square, model and corresponding variable is also significant.

If the magnitude of f-value is too large then associations between variables is long term and if magnitude comes close to 3.84 but more than 3.84 then associations between variables is short term.

Schwarz SC > Akaike A/C = Schwarz criteria is used

Schwarz SC < Akaike A/C = Akaike criteria is used

ET = 1 – R square

Vector Autoregression Estimates
Date: 11/27/11 Time: 18:36
Sample (adjusted): 3 60
Included observations: 58 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	MARKET_1	MARKET_2
MARKET_1(-1)	1.206822	-0.274478
	(0.13608)	(0.60479)
	[8.86817]	[-0.45384]
MARKET_1(-2)	-0.298095	0.360874
	(0.12935)	(0.57488)
	[-2.30448]	[0.62773]
MARKET_2(-1)	-0.023495	1.522074
	(0.02828)	(0.12569)
	[-0.83079]	[12.1101]
MARKET_2(-2)	0.039556	-0.537739
	(0.02972)	(0.13208)
	[1.33102]	[-4.07139]
C	309.2539	227.9989
	(80.2309)	(356.566)
	[3.85455]	[0.63943]
R-squared	0.998789	0.999194
Adj. R-squared	0.998697	0.999133
Sum sq. resids	384861.4	7601531.
S.E. equation	85.21464	378.7151
F-statistic	10926.42	16417.92
Log likelihood	-337.5041	-424.0175
Akaike AIC	11.81049	14.79371
Schwarz SC	11.98811	14.97133
Mean dependent	7923.605	33201.57
S.D. dependent	2361.070	12859.95
Determinant resid covariance (dof adj.)		1.02E+09
Determinant resid covariance		8.54E+08
Log likelihood		-760.9909
Akaike information criterion		26.58589
Schwarz criterion		26.94114

Findings:

1) Variables are showed in absolute form. Because there is no sign of delta (△) with variables. (i.e market_1 & market_2)

2) Market_1 is taken to be consideration b/c our DV is market_1

3) Model:

$X = \alpha + \beta 1lag1X + \beta 2lag2X + \beta 3lag1Y + \beta 4lag2Y + ET$

	M1 = α +	β1lag1M1 +	β2lag2M1 +	β3lag1M2 +	β4lag2M2 +	ET
Alpha value	309.25	1.20	-0.29	-0.02	0.03	
t-value	(3.85)	(8.86)	(-2.3)	(-0.83)	(1.33)	

4) Lag1M1, lag2M2 are significant and lag1M2, lag2M2 are not significant.

5) Previous year lag1 & lag2 of M1 are predicting DV but lag1 & lag2 of M2 are not predicting DV that's why it come in the category of AR not VAR. Because DV is regressing with its previous year lags but not other variable lags.
6) Intercept is significant. That means, if all IV come zero then there will be some value of DV. Mean intercept has some value.
7) In SPSS-13 Rho was 0.94. In e-views 1.2. (So 0.94 ~ 1.2)
8) Adjusted R square = 0.99. That means, if DV is 99% predicted by the model.
9) F-value = 10926.42. That means, association between variables is long term. In other words, association of lag1M1 and lag2M1 with M1 is long term.
10) Magnitude of Schwarz SC > magnitude of Akaike A/C. Which means that Schwarz criteria is applied.

Unidirectional/Bidirectional Causality:

Two variables can predict in not time series. Regression is used weather it is bi-variant or multi-variant.

Two variables can cause in time series. A variable which causes, should be in previous lag and a variable which is caused, should be in current lag.

Xt-1 effecting Yt Unidirectional

Xt-1 effecting Yt & Yt-1 effecting Xt Bidirectional

1) For unidirectional, Granger causality test is used.
2) For bidirectional, Cross correlation test is used.

1) The purpose of Granger causality test is to investigate unidirectional causality.
2) The purpose of Cross correlation test is to investigate bidirectional causality.

Unidirectional Causality:

Granger Causality Test: (testing unidirectional causality)

Model will be in OLS frame work and intercept is not present in model.

$Yt = \beta Xt\text{-}1 + ET$ or

$Yt = \beta Xt\text{-}2 + ET$ or

$Yt = \beta Xt\text{-}3 + ET$

ET is present in every model:

It is not an ARLD model because lag variable is of IV and not of DV. So it is OLS frame work.

Why intercept is not present in model?

Reason:

If IV = 0 then DV = intercept.

Necessary condition = Cause = Cloud

Sufficient condition = Effect = Rain

1) Variable that addressed first = IV and variable that addressed after that = DV

2) Cause addressed first and effect addressed after that.

Without cause, effect is impossible whereas without IV, DV is zero.

So if IV = zero then DV = zero because there is no effect without cause.

There is no possibility of presence of intercept, if IV is zero.

Example:

In e-views (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1. (run)

Quick, group statistics, granger causality test.

Put market_1 space market_2

Select lag 1

1) Acceptance of Null hypothesis is more than 5% of p-value.

2) Acceptance of alternate hypothesis is less than 5% of p-value.

Pairwise Granger Causality Tests
Date: 12/04/11 Time: 17:16
Sample: 1 60
Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
MARKET_2 does not Granger Cause MARKET_1	59	6.08897	0.0167
MARKET_1 does not Granger Cause MARKET_2		0.43324	0.5131

Findings:

1) Null hypothesis:

   a) Market_2 does not Granger Cause market_1 (reject)

   b) Market_1 does not Granger Cause market_2 (accept)

2) Alternate hypothesis:

   a) Market_2 does Granger Cause market_1 (accept)

   b) Market_1 does Granger Cause market_2 (reject)

3) Result:

   a) Lag 1 of market_2 is significantly cause to market_1

   b) Lag 1 of market_1 is not significantly cause to market_2

4) Lag = sample (N) – Obs

   Lag = 60 -59

   Lag = 1 (therefore, Granger Causality is investigated for lag 1)

   Test took lag 1 of IV.

5) Model:

   $Yt = \beta Xt\text{-}1 + ET$

   $M1 = \beta M2t\text{-}1 + ET$

   f-value: (6.088)

Expected Question in Exam:

$Yt = \beta Xt\text{-}1 + ET$

= (f=6.088)

= N=60

= N of IV = 58

Granger Causality is investigated in which lag?

Answer: 60 – 58 = lag 2

Bidirectional Causality:

Partial auto regression and simple partial auto regression investigate non-stationarity in data.

They also tell that weather we should apply cross correlation or not.

In unidirectional causality:

X predicts Y (forward biased or reverse biased) irreversible reaction

Y predicts X (forward biased or reverse biased) irreversible reaction

Both are not predicting each other at a same time.

Unidirectional investigate 1 direction in 1 time.

In bidirectional causality:

X predicts Y and Y predicts X (reversible reaction) both are predicting each other at a same time

Bidirectional investigate 2 directions at a time.

J-BOX Q-Statistics method for PAC & SPAC:

Durbin Watson shows an autocorrelation in time series data.

H-test shows an autocorrelation in panel data.

J-box Q-statistics is used to investigate partial auto regression and simple partial auto regression in time series data and panel data.

Durbin Watson just tells that auto correlation is present or not but J-box Q-statistics tells that which lag is predicting current observation.

J-box Q-statistics tells categorically each observation about each lag.

Xt-1 is predicting Y (auto correlation)

Frame work of J-box Q-statistics of auto correlation.

X	Xt-1	Xt-2	Xt-3	Xt-4
1	2	3	4	5
2	3	4	5	-
3	4	5	-	-
4	5	-	-	-
5	-	-	-	-

Number of lags = Number of observations – 1

Q. Should we include intercept in model?

If all IV is zero then DV and α will be 1st value of DV i.e. (1)

Therefore intercept should be present in model.

Co-linearity is investigated among IV and Auto correlation is investigated in DV only due to intercept. (α)

1st value of DV can't be zero because it is scale data. Zero value does matter in categorical data but there will be no observation of zero value in scale data.

$Xt = \alpha + \beta 1Xt\text{-}1 + \beta 2Xt\text{-}2 + \ldots\ldots\ldots + \beta nXt\text{-}n + ET$

Here magnitude of α is the 1st value of DV.

There are two events come in J-box Q-statistics those indicate the stationarity in data and auto correlation.

1) Cutoff event = auto correlation varies immediately. It shows temporary relation.
2) Dicing event = auto correlation decrease gradually in lags. Is shows permanent relation.

These events should be present b/c we want auto correlation.

ADF and PAC investigate non-stationarity in data.

Example:

In SPSS-17 (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1.

(run)

Analyze, forecasting, autocorrelation

Click on autocorrelation and partial autocorrelation.

If data was panel: Then we select option & do click on (display autocorrelation at periodic lag)

If two time series is present and autocorrelation is present b/w those for same lag so chances of presence of bidirectional causality.

For market 1:

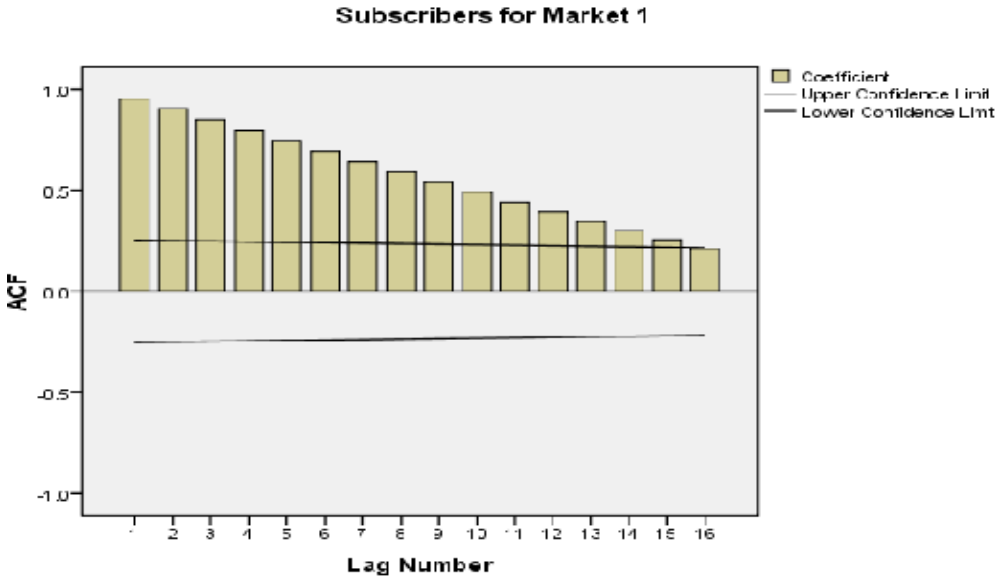
Autocorrelations

Series:Subscribers for Market 1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.954	.126	57.432	1	.000
2	.905	.125	110.001	2	.000
3	.853	.124	157.451	3	.000
4	.799	.123	199.852	4	.000
5	.745	.122	237.481	5	.000
6	.694	.120	270.372	6	.000
7	.642	.119	299.336	7	.000
8	.592	.118	324.724	8	.000
9	.541	.117	346.078	9	.000
10	.492	.116	364.061	10	.000
11	.443	.115	378.933	11	.000
12	.394	.114	390.975	12	.000
13	.347	.112	400.489	13	.000
14	.301	.111	407.799	14	.000
15	.254	.110	413.148	15	.000
16	.212	.109	416.935	16	.000

a. The underlying process assumed is independence (white noise).

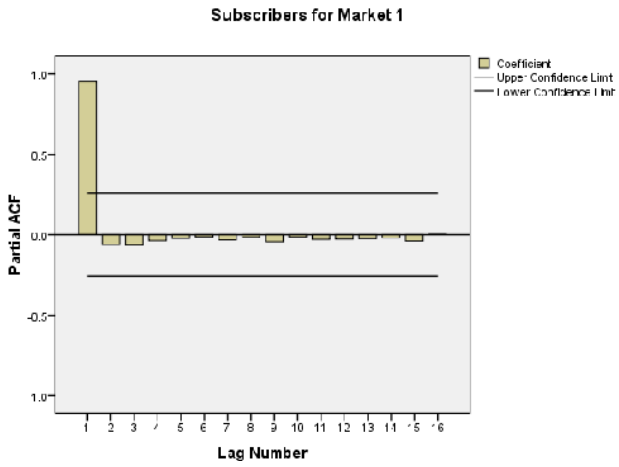
b. Based on the asymptotic chi-square approximation.



Partial Autocorrelations

Series:Subscribers for Market 1

Lag	Partial Autocorrelation	Std. Error
1	.954	.129
2	-.082	.129
3	-.085	.129
4	-.037	.129
5	-.023	.129
6	-.012	.129
7	-.033	.129
8	-.015	.129
9	-.043	.129
10	-.014	.129
11	-.029	.129
12	-.028	.129
13	-.025	.129
14	-.021	.129
15	-.039	.129
16	.002	.129



For market 2

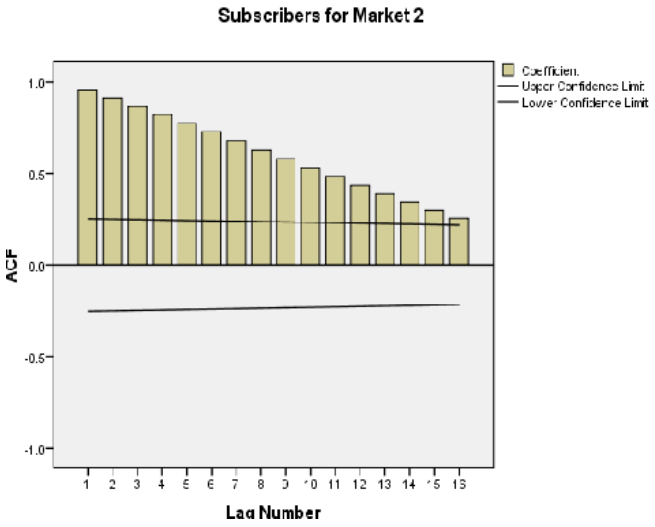
Autocorrelations

Series:Subscribers for Market 2

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.955	.125	57.485	1	.000
2	.911	.125	110.589	2	.000
3	.857	.124	159.728	3	.000
4	.822	.123	204.575	4	.000
5	.775	.122	245.274	5	.000
6	.729	.120	281.890	6	.000
7	.681	.119	314.395	7	.000
8	.631	.118	342.833	8	.000
9	.581	.117	367.458	9	.000
10	.532	.116	388.492	10	.000
11	.484	.115	405.305	11	.000
12	.437	.114	421.117	12	.000
13	.391	.112	433.228	13	.000
14	.345	.111	442.858	14	.000
15	.299	.110	450.299	15	.000
16	.255	.109	455.743	16	.000

a. The underlying process assumed is independence (white noise).

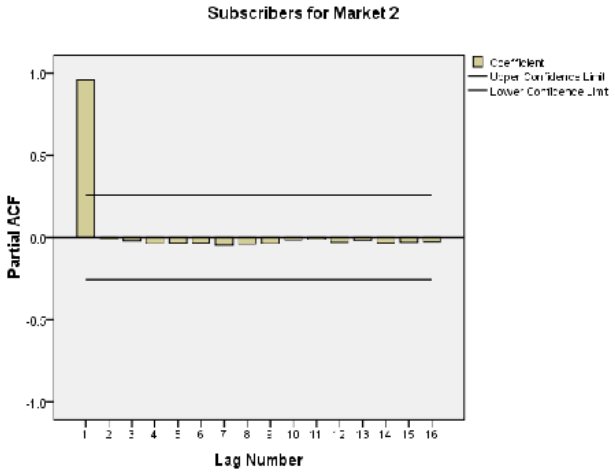
b. Based on the asymptotic chi-square approximation.



Partial Autocorrelations

Series:Subscribers for Market 2

Lag	Partial Autocorrelation	Std. Error
1	.955	.129
2	-.011	.129
3	-.021	.129
4	-.038	.129
5	-.034	.129
6	-.034	.129
7	-.048	.129
8	-.042	.129
9	-.037	.129
10	-.017	.129
11	-.012	.129
12	-.030	.129
13	-.018	.129
14	-.031	.129
15	-.029	.129
16	-.027	.129



Findings:

1) In Partial Autocorrelation cutoff event is present in charts of market 1 & 2 so stationarity is present in both markets.
2) Market 1 and market 2 are significant for 16 lags.
3) Significant lags are same for market 1 & 2, graphs are also same (i.e. cutoff). So bidirectional causality is investigated and we can apply cross correlation.
4) Cutoff event is indicating temporary relation between variables.

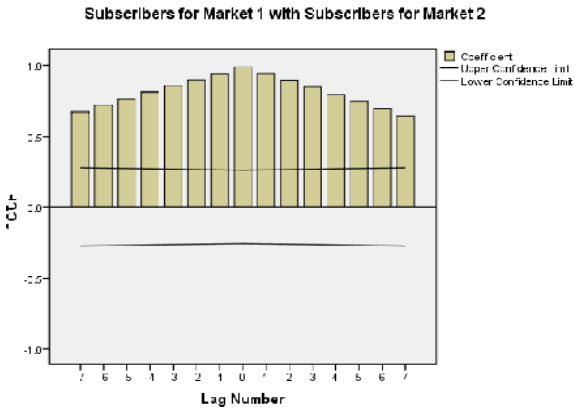
Cross correlation:

Cross Correlations

Series Pair:Subscribers for Market 1 with Subscribers for Market 2

Lag	Cross Correlation	Std. Error[a]
-7	.673	.137
-6	.719	.136
-5	.765	.135
-4	.810	.134
-3	.855	.132
-2	.899	.131
-1	.943	.130
0	.988	.129
1	.944	.130
2	.897	.131
3	.847	.132
4	.796	.134
5	.746	.135
6	.695	.136
7	.644	.137

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.



Findings:

1) Column of cross correlation is showing cross correlation between both variables i.e. market1 & 2
2) Maximum prediction is on lag 0
3) There are no cross correlation b/w variables.
4) White noise = ET

   ET is present any one of the series. That can be market 1 or 2.

5) It always comes with chart that present of white noise in data which proves that it is not an axiom.
6) In future prediction and forecasting, cause will be decrease in future.
7) Lag 5 of market 1 is predicting 76.5% of current variable of market 2 and lag 5 of market 2 is predicting 76.5% of current variable of market 1.

Co-Integration:

Co-integration means co-movement.

If co-integration is significant so it means that both series are exclusively move in a same direction with each other. In other words co-movement is present in both series.

Mackinnon critical value 5% is P-value.

Types

1) Johnson co-integration
    a) Bi-variant
    b) Multi-variant
2) VAR based co-integration. (application of VAR)
3) ECM based co-integration. (application of ECM)

Bi-Variant:

If we investigate test in pair form then we will use Bi-Variant.

If T-statistics > Mackinnon Critical value

Presence of non-stationarity

Log-likely hood is a coefficient of co-integration.

Example:

In e-views (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1. (run)

Quick, group statistics, co-integration

Write market_1 (space) market_2

At MHM Mackinnon critical value is 5%

We have to select Osterwald Lenun for random basis. Means 1% and 5%

Johansen Cointegration Test

Date: 12/18/11 Time: 18:10
Sample (adjusted): 3 60
Included observations: 58 after adjustments
Trend assumption: Linear deterministic trend
Series: MARKET_1 MARKET_2
Lags interval (in first differences): 1 to 1

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.141314	8.851337	15.41	20.04
At most 1	0.000258	0.014952	3.76	6.65

Trace test indicates no cointegration at both 5% and 1% levels
*(**) denotes rejection of the hypothesis at the 5%(1%) level

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None	0.141314	8.836385	14.07	18.63
At most 1	0.000258	0.014952	3.76	6.65

Max-eigenvalue test indicates no cointegration at both 5% and 1% levels
*(**) denotes rejection of the hypothesis at the 5%(1%) level

Findings:

1) Trace statistics = t-statistics = t-value < Mackinnon critical value (at None)

   There is no co-integration

2) Max-Eigen statistics = confirmation test

   t-value < Mackinnon critical value (at None)

   There is no co-integration

3) Both are not co moving & non-stationarity is present in different order

4) Chance of co-integration at different order is more than same order.

<u>Multi-Variant:</u>

If we investigate test in joint form then we will use Multi-Variant.

If T-statistics > Mackinnon Critical value

Presence of non-stationarity

Log-likely hood is a coefficient of co-integration.

Example:

In e-views (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1. (run)

Quick, group statistics, co-integration

Write market_1 (space) market_2 (space) market_3

Now we will select MHM for Mackinnon critical value 5%

Johansen Cointegration Test

Date: 12/18/11 Time: 18:13
Sample (adjusted): 3 60
Included observations: 58 after adjustments
Trend assumption: Linear deterministic trend
Series: MARKET_1 MARKET_2 MARKET_3
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.257573	26.79046	29.79707	0.1069
At most 1	0.137464	9.516281	15.49471	0.3198
At most 2	0.016065	0.939351	3.841466	0.3324

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None	0.257573	17.27418	21.13162	0.1595
At most 1	0.137464	8.576930	14.26460	0.3231
At most 2	0.016065	0.939351	3.841466	0.3324

Max-eigenvalue test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Findings:

1) At None* = co-integration b/w market_1 and market_2

At most 1* = co-integration b/w market_1 and market_3

At most 2* = co-integration b/w market_2 and market_3

There is no star at any hypothesized. So there is no co-integration b/w variables.

2) Trace statistics = t-statistics = t-value < Mackinnon critical value (at all level)

   So co-integration has not been found

3) Max-Eigen statistics = confirmation test

   t-value < Mackinnon critical value (at all level)

   There is no co-integration

Q. why do we call error is white noise?

Error means resistance or noise and white color is the combination of all color. So when all errors are combined so it becomes white noise. In other words "merger and aggregation of error is called white noise".

Example:

Table1: Augmented Dickey-Fuller (ADF) Unit Root Test:

VARIABLES / SERIES	ADF TEST WITH INTERCEPT & TREND AT SAME ORDER			ADF TEST WITH INTERCEPT & TREND AT 1ST DIFFERENCE		
	ADF COEF	T-STAT	P-VALUE	ADF COEF	T-STAT	P-VALUE
Stock Prices of KSE 100-Index	-0.000226	-0.53938	0.8810	-0.864638	-50.54710	0.0001
Stock Prices of BSE Sensex	-4.45 E-05	-0.08453	0.9943	-0.925497	-54.18894	0.0000
Stock Prices of DSE Composite	-0.001420	-0.82841	0.8101	-1.080524	-24.39047	0.0000
Stock Prices of Nepal SExchange	-0.000172	-0.33944	0.9188	-0.940423	-56.62769	0.0000
*Mackinnon Critical Values (5008 Observations)			1% - 3.43 5% - 2.86 10% - 2.56			

Findings:

1) Data of four stock exchanges has been taken.

   Karachi, Bombay, Dhaka, and Nepal

2) T-statistics > Mackinnon Critical value

   Presence of non-stationarity.

3) Now we will investigate co-integration among variables

Table 2: Johansen Multivariate Co-integration Test among the Stock Prices of KSE 100, BSE Sensex, DSE Composite, and Nepal SE indices

VARIABLES / SERIES	LOG-LIKELIHOOD	T-STATISTICS
Stock Prices of KSE 100, BSE Sensex, DSE, and NSE Indices	-82748.16	50.507080
Mackinnon Critical Values (5008 Observations)	1% 54.46 5% 47.21	

Table 3: Johansen Bi-variate Co-integration Test between the Stock Prices of KSE 100-Index & rest of equity markets on 1 on 1 basis

VARIABLES / SERIES	LOG-LIKELIHOOD	T-STATISTICS
Stock Prices of KSE 100-Index & Bombay Stock Exchange Sensex	-42408.20	11.374460
Stock Prices of KSE 100-Index & Dhaka Stock Exchange Composite	-42899.62	15.620252
Stock Prices of KSE 100-Index & Nepal Stock Exchange	-38299.68	11.228000
Mackinnon Critical Values (5008 Observations)	1% 20.04 5% 15.41	

Findings:

1) Co-integration has been found at 5% among all variables in multi-variate co-integration test

2) Co-integration has been found at 5% b/w KSE & Dhaka Stock Exchange in Bi-variate co-integration test

3) There is no co-integration b/w KSE & Bombay, KSE & Nepal in Bi-variate

4) Log-Likelihood is a co-efficient of co-integration

   The equations of KSE & Dhaka Stock Exchange are associated with co-efficient (-42899.62)

Error Correction Model: (ECM) or Granger Representation Theorem:

It explains the correction of error in model of association of variables.

Time series data (at least 2)
↓
Non-stationarity at zero/same order & stationarity at not zero order
↓
Exogenous regressor is significant for both series
↓
ECM

If we have at least 2 time series data then we check non-stationarity at zero/same order and stationarity at not zero order (can be at 1st difference = 1st lag) and exogenous regressor is also significant then we will apply ECM.

Exogenous regressor indicates clearly that weather we will apply ECM or not.

Another school of thought is: non-stationarity in time series is an issue, so it should be fixed. So fixing of non-stationarity is called stationarity.

Example:

In e-views (Open file) C/program files/SPSSInc/statistics17/samples/English/broadband_1. (run)

Quick, VAR, click on Vector error correct

Write market_1 (space) market_2

If f-statistic is significant (means more than 3.84) then error has been corrected and findings have come after correction. But if f-statistic is not significant then we will not accept values of 2nd chart.

t-value = trace value = significant when t-value>1.5 (weather +ve or –ve)

CointEq1 = equation of market_1 current lag

[ ] = t-statistics value

( ) = standard error

Vector Error Correction Estimates

Vector Error Correction Estimates
Date: 12/25/11 Time: 16:55
Sample (adjusted): 4 60
Included observations: 57 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
MARKET_1(-1)	1.000000
MARKET_2(-1)	-0.173134
	(0.00833)
	[-20.7776]
C	-2173.939

Error Correction:	D(MARKET_1)	D(MARKET_2)
CointEq1	-0.104501	0.130622
	(0.03209)	(0.14496)
	[-3.25635]	[0.90112]
D(MARKET_1(-1))	0.234602	-0.283161
	(0.13011)	(0.58770)
	[1.80311]	[-0.48181]
D(MARKET_1(-2))	0.222309	-0.579494
	(0.13597)	(0.61418)
	[1.63497]	[-0.94353]
D(MARKET_2(-1))	-0.039881	0.459989
	(0.03135)	(0.14162)
	[-1.27199]	[3.24800]
D(MARKET_2(-2))	-0.014173	0.180992
	(0.03342)	(0.15095)
	[-0.42410]	[1.19902]
C	108.8391	398.6571
	(36.8868)	(166.616)
	[2.95063]	[2.39267]

R-squared	0.284784	0.286744
Adj. R-squared	0.214665	0.216817
Sum sq. resids	351628.0	7174223.
S.E. equation	83.03413	375.0614
F-statistic	4.061421	4.100613
Log likelihood	-329.6069	-415.5537
Akaike AIC	11.77568	14.79136
Schwarz SC	11.99074	15.00642
Mean dependent	134.2944	726.9841
S.D. dependent	93.69772	423.8095

Determinant resid covariance (dof adj.)	9.54E+08
Determinant resid covariance	7.63E+08
Log likelihood	-744.6782
Akaike information criterion	26.62029
Schwarz criterion	27.12209

Findings:

1) First part shows a model without correction.
2) Second part indicates the values after correction.
3) F-statistic is significant for both variable (i.e. market_1 & market_2)
4) By t-value: 1st difference of market_1 is predicted by cointEg1, 1st difference of market_1 lag 1, 1st difference of market_1 lag 2. And intercept. Whereas it is not predicted by 1st difference of market_2 lag 1 and 1st difference of market_2 lag 2.
5) By t-value: 1st difference of market_2 is predicted by 1st difference of market_2 lag 1 and intercept. Whereas it is not predicted by cointEg1, 1st difference of market_1 lag 1, 1st difference of market_1 lag 2 and 1st difference of market_2 lag 2.

Endogenity:

It is measureable.

If any relevant variable did not include in model then it is called ignored variable and it is will be include in ET. If that ignored variable is co-linear with any IV so due to this ET will also co-lineant with that IV.

$Qd = \alpha - \beta1(P) + \beta2(Y) + ET$

Now if: ignored variable = quality =Q

Quality is measureable so its endogenity. And we will apply Heckman regression.

If we apply Hacksman regression and in comes insignificant then it means that error term comes from exogenity.

Exogenity:

It is not measureable.

If any relevant variable did not include in model then it is called ignored variable and it is will be include in ET. If that ignored variable is co-linear with any IV so due to this ET will also co-lineant with that IV.

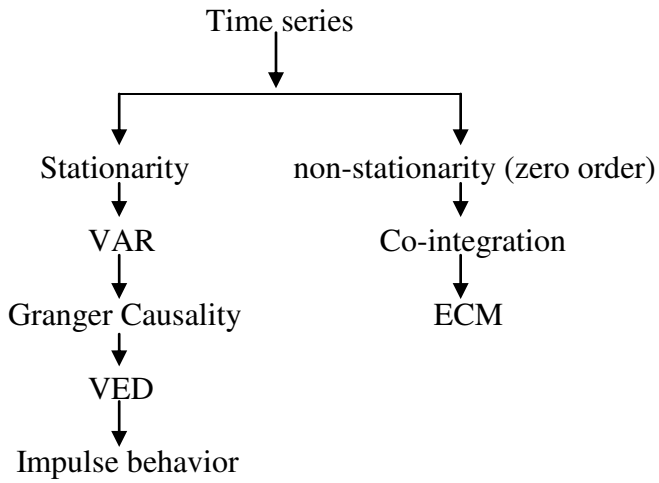
$Qs = \alpha - \beta 1(P) + \beta 2(quality) + ET$ (Production of fruit)

Now if: ignored variable = Sunlight

Sunlight is not measureable so its exogenity. And we can't apply any test on it. It can't be measureable b/c its variable is not measureable.

We can apply ECM only on endogenity.

Sum Up All Tests:



For Stationarity

1) **Stationarity:** If data is time series then we will apply ADF URT. If bumps are not present then stationarity is present. Weather it is on same order or 1st difference or 2nd difference.
2) **VAR:** If stationarity is present then we will apply VAR. VAR predicts, and it is significant when PAC tells cutoff & dicing.

   VAR: Tells previous lag of 1st series is predicting the 2nd series current lag and previous lag of 2nd series is predicting the1st.

   For one lag = AR and for may lags VAR (P) order
3) **Granger Causality:** If VAR is significant then chance of present of granger causality. It causes not predicts.
4) **VED:** If granger causality is present then we investigate VED (variance error decomposition) on a series which is caused.

It can tell can we minimize SD or can we minimize the difference b/w mean of ET – lag observation of ET.

VED: it tells that weather we can decompose variance of error term or not.

Variance = $(S.D)^2$ = (mean of ET – lag observation of ET)

5) **Impulse Analysis:** If VED is significant then we apply impulse behavior or impulse analysis.

   Any minor change is impulse behavior. Visible change = cutoff = bumps

<u>For Non-Stationarity</u>

1) **Non-stationarity:** If data is time series then we will apply ADF URT. If bums are present then non-stationarity is present.
2) **Co-integration:** If non-stationarity is present in 2 or more time series data then we investigate co-integration.

   If non-stationarity is on zero order then co-integration is confirmed 100%

   If non-stationarity is on same order then co-integration is confirmed 50%

   If non-stationarity is on different order then co-integration is confirmed 0%
3) **ECM:** Beside this if exogenous regressor is significant as well then we apply ECM also. The concept of unidirectional and ECM is given by granger so the other name of ECM is granger representation theorem.

Non-Time Series Econometrics:

Prediction and application of regression:

For prediction = OLS model is used & regression is used as econometric technique.

IV = fixed / same

DV = random

IV is same every time means prediction is not accurate and change every time.

$Y = \alpha + \beta X$

$3 = \alpha + \beta (2)$

$6 = \alpha + \beta (2)$

$14 = \alpha + \beta (2)$

Types of regression:

- **On the basis of linearity or non-linearity.**

  1) **Linear Regression**

     a) Simple linear regression (SLR)

     b) Multiple linear regression (MLR)

     c) GLM

     d) Logit (multi-nominal regression)

     e) Probit (dichotomus regression)

     f) Categorical regression/Optimal scaling

     g) Tobit

     h) Hecksman regression

     i) Partial regression

2) **Curvi-linear Regression**

   a) Exponential regression (equations are depend on number of derivatives taken)

   b) Quadratic regression (having 2 equations)

   c) Cubical regression (having 3 equations)

Exponential Regression:

$DV = \alpha + IV1^{\beta 1} + IV2^{\beta 2}$

1st derivative = 1st equation

2nd derivative = 2nd equation

3rd derivative = 3rd equation

- **On the basis of model:**

1) **OLS model (ordinary least square)**

   a) Simple linear regression (SLR)

   b) Multiple linear regression (MLR)

   c) GLM

   d) Logit (multi-nominal regression)

   e) Probit (dichotomus regression)

   f) Categorical regression/Optimal scaling

   g) Hecksman regression

   h) Exponential regression

   i) Quadratic regression

   j) Cubical regression

2) **SOLS model (Scaled ordinary least square)**

   a) Tobit

3) WLS (weighted least square)

a) Partial Regression

References:

[1] Greene, W. H. (2011). Econometric Analysis (7th Edition). Publisher: Prentice Hall; 7 edition, ISBN-13: 978-0131395381, ISBN-10: 0131395386

[2] Gujarati, D. N. (2009). Basic Econometrics (Int'l Ed). Publisher: McGraw-Hill Higher Education; 5 edition, ISBN-10: 0071276254, ISBN-13: 978-0071276252

[3] Kmenta, J. (1997). Elements of Econometrics: Second Edition. Publisher: University of Michigan Press; 2 Sub edition, ISBN-10: 0472108867, ISBN-13: 978-0472108862

[4] Koutsoyiannis, A. (2001). Theory of Econometrics. Publisher: Palgrave Macmillan Limited, 2nd edition. ISBN: 0333778227, 9780333778227

[5] Maddala, G. S. & Lahiri, K. (2009). Introduction to Econometrics, 4th Edition. Publisher: Wiley E-Text, ISBN: 978EUDTE00508

Synonyms Glossary:

Critical value = significant value

Spectral analysis = de-trending

Optimal scaling = linier scaling = categorical regression

Censored regression = propid

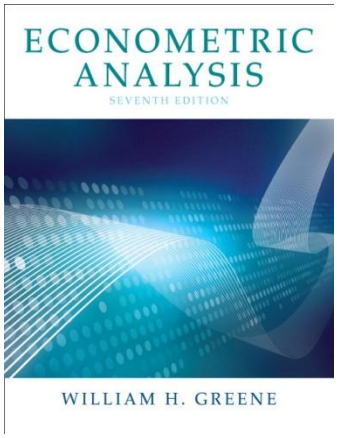
Curvi-linier regression = non-linier regression

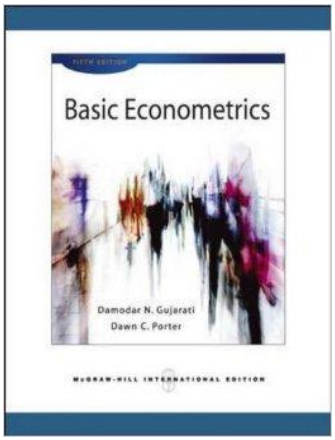
Non-parametric nominal data = categorical data

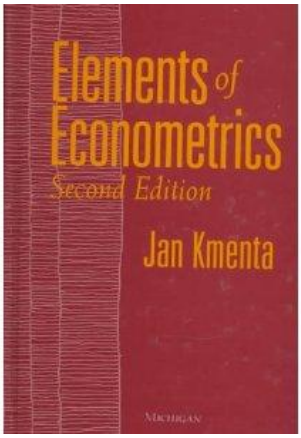
Non-stationarity = oscillation = bumps = shocks

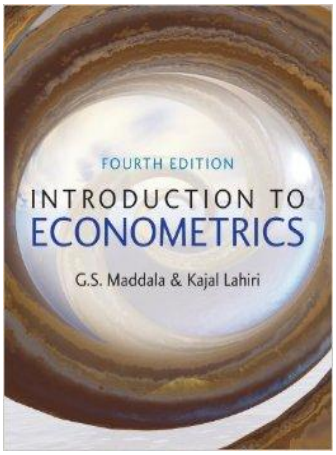
flyleaf - 3

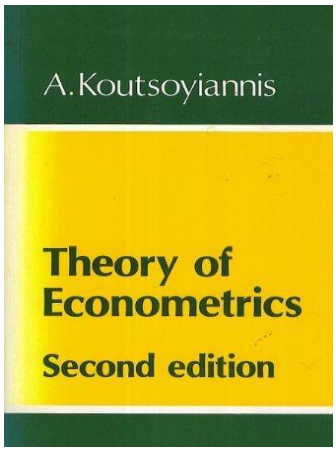
This page intentionally left blank











Few recommended books for more knowledge about Econometrics

Mohsin Hassan Alvi

Bachelor of Commerce
University of Karachi, Karachi, Pakistan

Master of Business Administration (Marketing & Finance)
Iqra University, Main Campus, Karachi, Pakistan